Central Limit Thereom Review

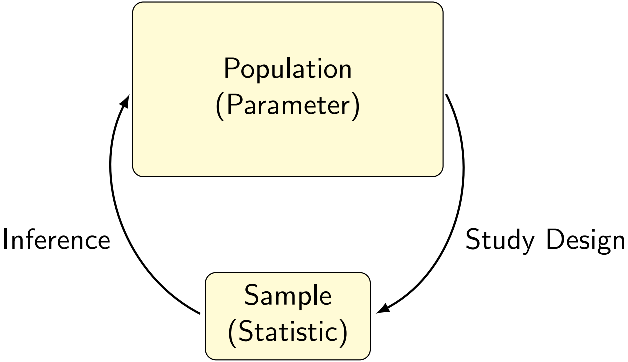
Grinnell College

March 24, 2025

Warm-up

- What is a sampling distribution and why is it useful?
- According to the Central Limit Theorem, what is the distribution of the sample mean?
- How do confidence intervals associate the percentiles of a sampling distribution with its standard error?
- What three elements impact the *width* of a confidence interval?

The Statistical Framework



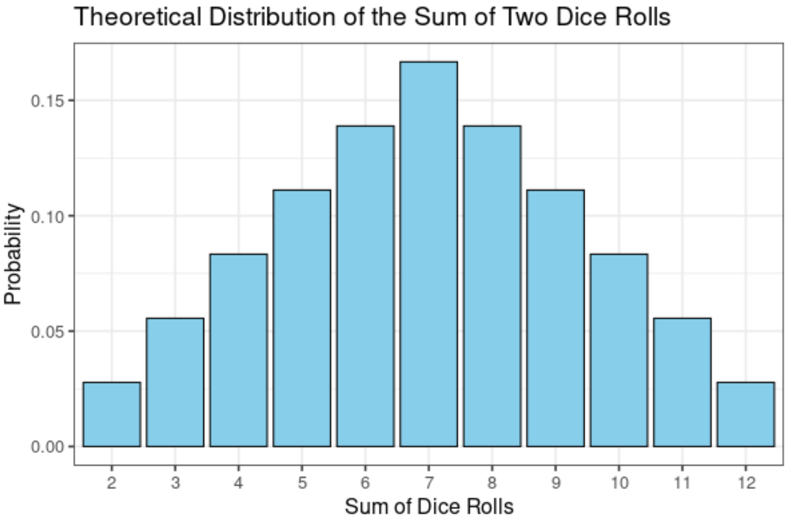
Distributions

Recall that a **distribution** tells us:

- What values
- How frequently

Most distributions are governed by **distributional parameters**: if we know these, we know everything we can about the data-generating process

Dice Sum	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$



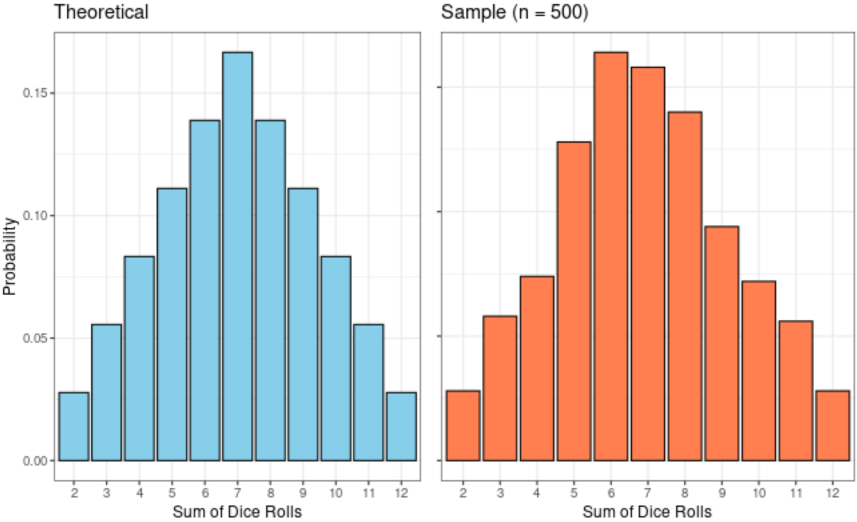
Random Samples

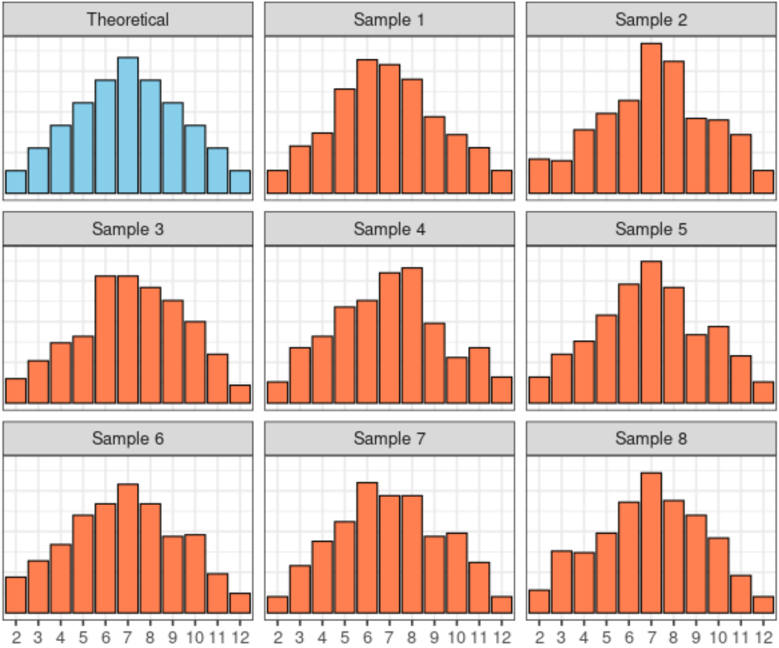
When we don't know the distributional parameters, we are instead required to take a sample from the population. If done correctly, this sample should be **representative**

The goal of any sample is to compute a statistic and perform **inference** on the unknown population parameter

Sampling is a **random process**, and this randomness will be reflected in the values of the statistics we are able to compute

Dice Sum	2	3	4	5	6	7	8	9	10	11	12
Probability	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$




Theoretical
Sample 1
Sample 2
Sample 3
Sample 4
Sample 5
Sample 6
Sample 7
Sample 8
2 3 4 5 6 7 8 9 10 11 12
2 3 4 5 6 7 8 9 10 11 12
2 3 4 5 6 7 8 9 10 11 12

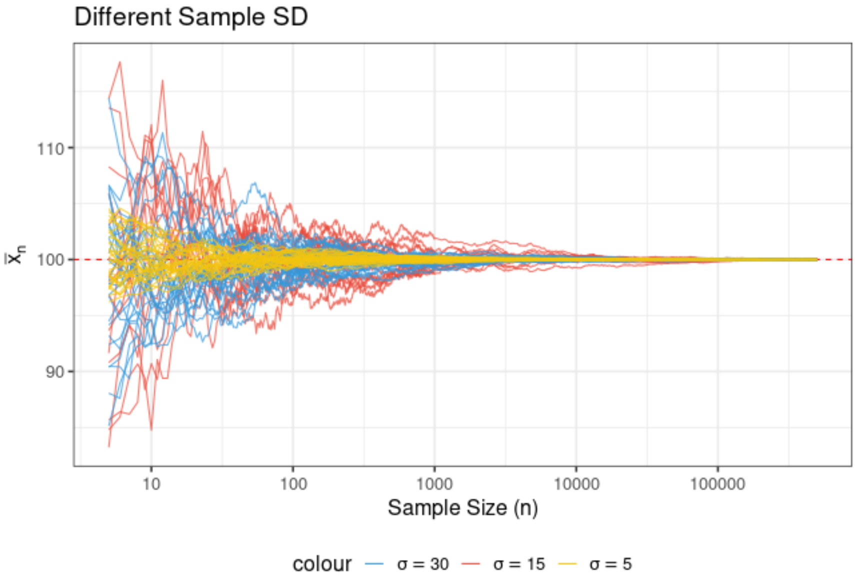
Sampling Distributions and CLT

The **sampling distribution** describes the distribution of a computed statistic. It tells us where our distribution should be centered and how much uncertainty exists around that center

The **Central Limit Theorem** states that for a population with mean μ and standard deviation σ, the sample mean will approximately follow a **normal distribution** with

$$\overline{X} \sim N\left(\mu,\ \frac{\sigma}{\sqrt{n}}\right)$$

We see that both the variance in the population and the sample size play a role in the amount of uncertainty in our estimate of $\overline{X}$



Different Sample SD
110
100
90
$\bar{x}_n$
10
100
1000
10000
100000
Sample Size (n)
colour
σ = 30
σ = 15
σ = 5

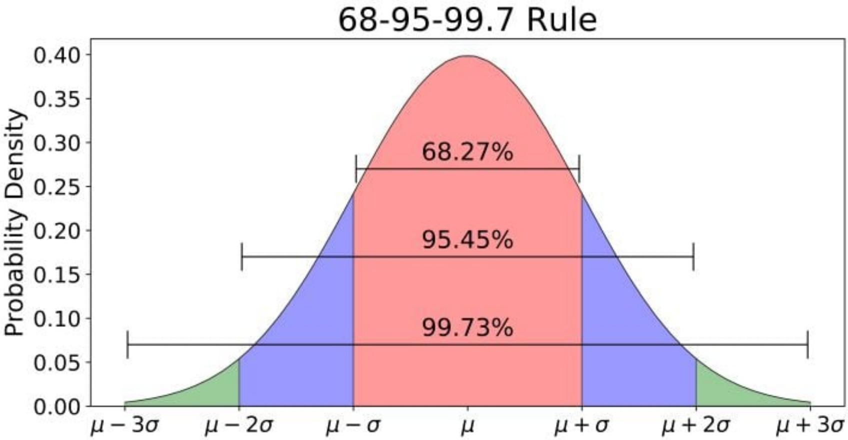
Confidence Intervals

Ultimately, our goal is to create a range of plausible values for the population parameter μ based on our sample

Typically, this is done with **percentiles**, with the range of our percentiles indicating our **confidence level**

For example, we may wish to construct an interval that has the property: "If I were to repeat the sampling process and generate an interval the same way, 95% of the time this interval will contain the true value for μ"

The Central Limit Theorem gives us a way to express this confidence in terms of standard error


68-95-99.7 Rule
Probability Density
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
68.27%
95.45%
99.73%
μ − 3σ
μ − 2σ
μ − σ
μ
μ + σ
μ + 2σ
μ + 3σ

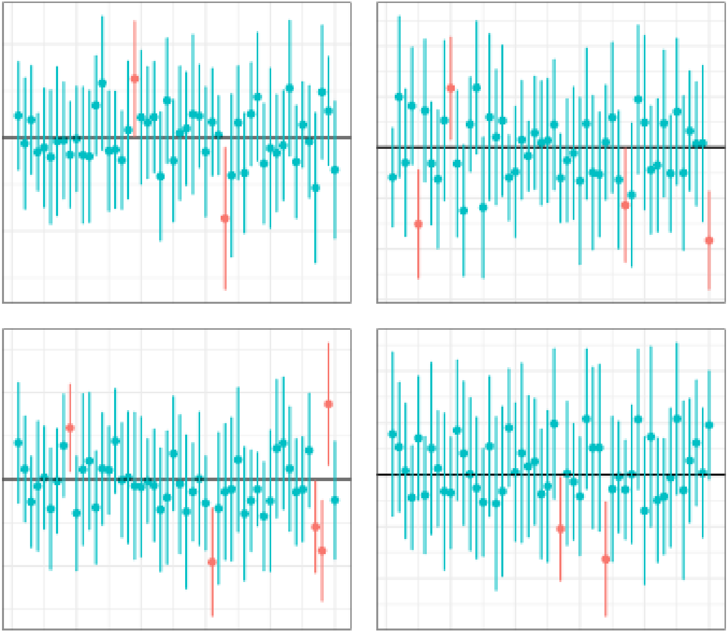
Confidence Intervals

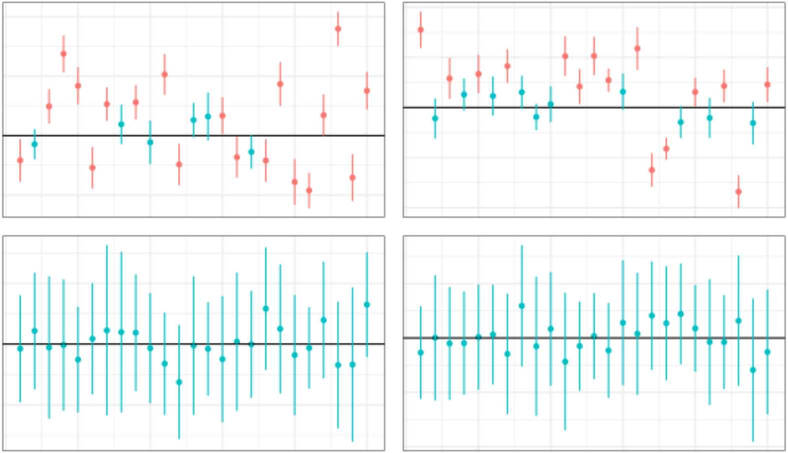
Specifically, we can express our confidence in the form of the interval

$$\overline{X} \pm C \times \frac{\sigma}{\sqrt{n}}$$

where $\sigma/\sqrt{n}$ is the standard deviation of our sampling distribution (i.e., the standard error)

The value C is called our **critical value**. While C, σ and n all determine the *width* of a confidence interval, only C mediates how much *confidence* we wish to have





Key Terms

- Sampling distribution
- Central Limit Theorem
- Normal distribution
- Percentiles
- Standard error
- Confidence level
- Critical values